
Interaction Homme Machine pour l'orientation spatiale des malvoyants

Said Boularouk^{1,2}, Didier Josselin^{2,1}, Eitan Altman^{3,1}

1. Laboratoire d'Informatique d'Avignon, UAPV

said.boularouk@univ-avignon.fr

2. UMR ESPACE 7300, UAPV, Avignon

didier.josselin@univ-avignon.fr

3. INRIA, Sophia Antipolis

eitan.altman@sophia.inria.fr

RÉSUMÉ. Pour aider les malvoyants à compenser la perte visuelle et à appréhender l'environnement urbain extérieur, nous proposons un outil interactif basé sur l'automatisation de la langue via une interaction homme-machine. Nous présentons un simulateur, développé en CSharp, pour aider les personnes malvoyantes à connaître les lieux urbains extraits de la base de données géographiques d'OpenStreetMap. Le principe est de transmettre les informations spatiales par un module de synthèse vocale et recevoir les requêtes de le malvoyant par un module de reconnaissance vocale.

ABSTRACT. To help blind people to compensate visual perception and to better understand their outdoor environment, we developed a method using human-computer interaction via Text-to-Speech. It helps visually impaired people to know surrounding places from OpenStreetMap data by hearing. The principle is to convey spatial information by voice synthesis and receive requests from blinded people by voice recognition.

MOTS-CLÉS : Synthèse vocale, reconnaissance vocale, aveugle, malvoyant, cartographie, sig, sémiologie sonore, bases de données, appréhension de l'espace, interaction homme machine, géolocalisation, GPS, Haversine

KEYWORDS: speech synthesis, speech recognition, blind, visually impaired, mapping, gis, semiotics sound, databases, apprehension of space, human computer interaction, geolocation, GPS, Haversine

1. Introduction

Selon l'Organisation Mondiale de la Santé, il existe près de 285 millions de personnes qui présentent une déficience visuelle, dont 45 millions sont malvoyants et 180 millions souffrent de handicap visuel. 90 % des personnes qui présentent une déficience visuelle vivent dans un pays à faible revenu, 82 % des malvoyants sont âgés de 50 ans et plus¹. Les études démontrent que l'immobilité des malvoyants est une conséquence directe de leur handicap visuel. Pour remédier à ce problème, l'appréhension de l'espace par un champ perceptif autre que la vision s'avère nécessaire (Josselin *et al.*, 2016).

Les malvoyants ont besoin d'aide humaine ou matériel pour se déplacer. D'un côté, pour éviter les obstacles, de l'autre pour s'orienter dans l'espace. L'identification des lieux avoisinants est primordiale pour tout déplacement sécurisé et confortable. Cette identification est difficile, voire impossible pour le malvoyant seul dans un environnement méconnu. L'appréhension de l'espace passe par la connaissance des composantes de l'environnement. Nous ne considérons dans cet article que les déplacements extérieurs et excluons du champ les déplacements *indoor*.

En milieu urbain extérieur, les éléments de l'environnement qui peuvent intéresser les malvoyants (lieux à atteindre) sont représentés par des points (magasins, lieux de culture et de loisir, centres de santé, services de sécurité, administrations, etc.). Chaque composante est identifiée par son nom, sa localisation GPS (et son adresse). Sur certaines cartes en ligne, les composantes contiennent les détails du service qu'elle offre et d'autres caractéristiques techniques et urbaines. Au vu des progrès de la cartographie et de l'apparition du GPS, la collecte des positions et les détails des composantes géographiques est une chose acquise, à une précision variable selon les conditions de collectes de l'information (nombre de satellites, zones enclavées, qualité du matériel, mouvement). Dans cette article, nous proposons une solution adaptée aux malvoyants, qui illustre comment l'automatisation de la parole peut aider à construire une représentation de l'espace relativement comparable à la représentation «cartographique».

Au Japon, l'entreprise Human Tech of Japan a développé un système de navigation pour les personnes malvoyants, en utilisant le GPS et le smartphone sous *android* (Kushwaha, Kushwaha, 2011). Le système utilise la synthèse vocale du téléphone (TTS ou Text To Speech) pour aider le malvoyant dans ses déplacements (Tenbrink *et al.*, 2013). En Corée du Sud, une équipe de Hanyang Cyber University, a développé un autre système simple pour smartphone sous *android* permettant la navigation vocale (Cha *et al.*, 2013). D'autres systèmes de locomotion pour malvoyants existent déjà, comme la canne électronique d'aide à la détection d'obstacles (Cbel, Visioprtonic, etc.). Citons également,

1. <http://www.who.int/fr/>

par exemple, le système d'aide à la navigation Okeneea² dédié aux malvoyants, développé en France.

L'objectif de cette recherche est de proposer une application dédiée aux malvoyants pour appréhender l'espace, basée sur la géolocalisation, la recherche de lieux dans un environnement proche et le dialogue homme machine. Nous présentons un simulateur géographique permettant à un utilisateur malvoyant de se positionner dans un milieu urbain et chercher les lieux qui l'entourent dans un rayon choisi, modifiable par simple commande vocale. L'utilisateur peut également sauvegarder ses lieux préférés dans une liste de favoris, dans la perspective de tracer un itinéraire dont les repères sont importants et connus. Le simulateur peut transmettre la distance, l'adresse et les détails de chaque lieu. Cet outil, à l'état de prototype, peut potentiellement être utilisé par le malvoyant pour préparer ses déplacements et, à terme, pendant son déplacement lui-même pour la recherche de trajets adaptés. Dans son état actuel, le prototype ne permet pas encore de calculer des itinéraires optimaux adaptés aux malvoyants.

2. Le malvoyant dans l'espace

A cause du handicap visuel et des difficultés de déplacement dus à la cécité, les connaissances spatiales d'une personne malvoyante restent limitées, mais elles sont de même nature que celles des voyants (Hatwell, 2006). L'approche écologique de Gibson (Baltenneck, 2010; Virole, 2009; Hatwell, 2006) propose de situer la cognition au croisement de l'humain et de l'environnement dans lequel il évolue, en fonction de ce qu'il lui permet (ou non) de faire. Cette notion est très pertinente dans le cadre du déplacement d'une personne malvoyante, car elle nous apprend que la lecture de l'environnement est en partie individuelle. La théorie de Gibson a abouti à deux concepts principaux : le champ optique ambiant et les *affordances*. Le principe d'affordance désigne à la fois toutes les possibilités d'action sur un objet et la capacité d'un objet à suggérer sa propre utilisation. William Gaver considère l'affordance comme l'ensemble des propriétés de l'environnement offertes à un individu pour accomplir certaines actions (Gaver, 1992). Quand une personne malvoyante ne trouve pas dans le stimulus visuel l'information nécessaire pour se déplacer, parce que son champ de vision est déficient ou quasiment inexistant, elle se concentre sur les affordances reçues à travers les autres champs de perception.

Le concept d'affordance montre en effet «*la façon dont les organismes se couplent avec leurs environnements, (.../...) et décrit une relation structurelle entre la perception et l'offre de sens de l'objet*» (Virole, 2009). L'extension du concept des affordances au langage nous permet de comprendre que la per-

2. <http://www.okeenea.com/actualites/decouvrez-kapteo-une-revolution-dans-le-quotidien-des-personnes-deficientes-visuelles/>

ception est fondamentalement orientée vers la recherche de signification. Comprendre un mot est avant tout comprendre le sens qu'il renvoie. De même, un bruit entendu est signifiant dès lors qu'il correspond à une adaptation de l'organisme à son environnement, "*langage d'orientation naturaliste*" (Virole, 2009). D'ailleurs, il n'est pas nécessaire de faire appel à des analyses approfondies du fait linguistique pour observer que chez l'enfant, les mots sont appris parce qu'ils sont d'abord associés à des situations concrètes. Le support de la perception n'a pas vraiment d'importance si le «sens» de l'objet ciblé ou de l'action voulue est compris par le récepteur.

Certaines aides techniques (GPS, balises sonores...) peuvent compenser la déficience visuelle pour l'interprétation de l'espace. Elles sont utilisées différemment selon le type du handicap visuel de chaque individu (CFPSAA, 2010). L'automatisation de la parole peut jouer le rôle de transmetteur de l'information spatiale extraite d'une carte géographique. Grâce à la perception auditive, le malvoyant peut connaître l'emplacement et la distance des lieux et leur orientation relative.

Une enquête réalisée auprès des malvoyants par l'Institute of Gerontology et AIST au Japon (Miura *et al.*, 2011) a montré que les meilleurs moyens de réception des informations spatiales et de l'environnement durant leur mobilité sont les suivants : 48,8 % des participants malvoyants préfèrent la parole, 34,1 % choisissent le *pitch*³ et l'*harmonie sonore*⁴ et 31,7 % préfèrent les vibrations. L'étude ne montre pas de différence particulière entre les choix des participants malvoyants et les participants complètement aveugles. Cela signifie, très logiquement, que la perception auditive est le mode privilégié et commun chez les malvoyants pour compenser la perte visuelle pour l'orientation et l'appréhension de l'espace.

De Saussure (De Saussure, 1997) définit le signe comme signifié (sens) et signifiant (aspect physique du sens). La réalité d'une chose et sa signification (signe) est le résultat d'une convention entre les individus au sein d'une communauté. Cette convention est utile dans le domaine de la linguistique informatique et dans toute représentation graphique (exemple : légende d'une carte). La *représentation graphique* est une forme de codification d'information en signes graphiques, le *langage verbal* est aussi une forme de codification en signes sonores. La codification est tout simplement une transcription d'une information connue par l'intermédiaire d'un système de signes quelconques. Alors, pourquoi ne pas codifier l'information spatiale en signes sonores, plus spécifiquement dans notre cas, en texte parlé de synthèse ?

3. Ton ou hauteur d'une note, en lien avec sa fréquence ; plus le pitch augmente plus la note est aiguë.

4. En musique, l'harmonie sonore peut être considérée à la fois d'un point de vue général et théorique, mais également via l'écoute : un accord harmonieux est agréable à entendre.

3. Système d'Information Géographique vocal

Plusieurs arguments militent en faveur de l'interaction entre le sonore et le vocal. Le développement des technologies de la numérisation sonore arrive aujourd'hui à un niveau de développement très important. Le langage verbal, forme de codification en signes sonores, est le code de communication privilégié chez l'humain. Le développement de la reconnaissance et de la synthèse vocale par ordinateur accélère la recherche autour de l'interaction homme-machine. Enfin, les données vectorielles sont stockées dans des bases que l'on peut interroger et fouiller programmatiquement avec un temps de réponse très court, ce qui peut compenser la dimension spatiale de la vue.

On mesure l'efficacité d'une représentation par la facilité que rencontre le lecteur à la lecture d'un dessin. Cette facilité peut se mesurer par la différence entre le temps de perception d'une construction efficace et la perception d'une construction inefficace. Le sens auditif, mode de perception préféré des malvoyants, peut être une alternative pour recevoir l'information. En fait, la différence majeure entre la perception visuelle et la perception auditive est que la première est *spatiale*, tandis que la seconde est *fréquentielle* –la fréquence étant directement liée à sa période ($f = 1/T$)– et donc *temporelle*. Les deux sens participent de manière complémentaire à une perception de courte période, voire synoptique, et une perception plus chronologique donc séquentielle. Par ailleurs, le système nerveux central a la possibilité d'extraire du signal transmis par le nerf optique, des informations non spatiales comme la couleur, la texture ou la fréquence. De même, l'oreille a la possibilité d'extraire des signaux temporels et fréquentiels transmis par le nerf auditif et des informations relatives à l'espace (Jouhaneau, 2013). Il est possible de passer d'une représentation graphique à une représentation sonore, dont l'objectif reste de véhiculer l'information géographique aux malvoyants. Certes, les règles de construction d'une représentation sonore sont différentes de celles d'une représentation graphique. Dans notre étude, nous séparons d'abord le contenu (l'information) du contenant (le moyen de transmission de l'information). Ce qui nous importe, c'est la validité et l'efficacité de la transcription du contenu, c'est à dire la rapidité et l'intelligibilité de sa perception.

4. Un simulateur cartographique vocale

Sur une carte, on trouve des primitives géométriques (points, lignes et zones) qui sont indépendantes de la dimension et de la constitution de la tâche qui les rend visible (Bertin, 1973). Par exemple, un nœud est l'un des éléments de certains modèles de données (e.g. OpenStreetMap). Mais sa représentation, en vue de faire passer un message ou de le présenter comme un objet particulier (point localisé, intersection de routes d'un réseau, hub diffusant les flux...), peut varier sémiologiquement. Dans notre système, à part l'environnement de géométrie multiple dans lequel se trouvent les lieux, nous gérons

essentiellement un ensemble de points de l'espace, définis par leur latitude, leur longitude, leur identifiant et un vecteur contenant les détails de ce point (nom, type de services, adresses, propriétaire...). Ces données nous intéressent le plus, car nous souhaitons les transmettre aux malvoyants afin qu'ils appréhendent leur environnement et qu'ils puissent planifier leurs déplacements de façon sécurisée. Pour être compréhensible, ces informations doivent être transformées d'un langage formel en un langage humain, via un traitement de *parsing*, puis transmises au malvoyant dans sa propre langue, et ce de la façon la plus efficace. En complément, GoogleMap offre une API de géolocalisation inverse (Reverse geocoding) pour obtenir, à partir des coordonnées GPS, l'adresse du lieu. Notre système combine donc des données géographiques libres (OSM) et des APIs de différents domaines en ligne (OSM, GoogleMap) et fait interagir, dans un même environnement, un SIG en ligne et un système de Text-To-Speech (cf. Figure 1).

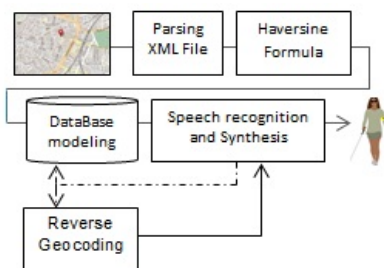


FIGURE 1. Schéma de principe et workflow

4.1. Extraction des données

Nous prenons la carte en ligne d'OSM comme source de données géographiques (Codescu *et al.*, 2011). Nous utilisons le module de reconnaissance et de synthèse vocale de Microsoft sous .NET pour l'interrogation et la communication des données pour le malvoyant.

Dans notre exemple, l'utilisateur est positionné à l'Université d'Avignon (marque rouge sur la Figure 2). À partir d'un fichier XML d'OSM (cf. Figure 3) et des attributs des objets localisés, nous pouvons extraire les positions de l'individu et des lieux environnants (boutiques, musée, soins de santé) qui sont représentés sur la carte, sous forme de *node* (nœud). Cela nous permet de calculer la distance entre chaque nœud et l'utilisateur malvoyant. En outre, nous pouvons obtenir toutes les informations importantes sur les endroits représentés par les *tags* (balises).

Le transfert des métadonnées du fichier OSM vers la base de données *Postgres* peut se faire via l'API *OSM2pgsql*. Il est important que le temps de réponse de

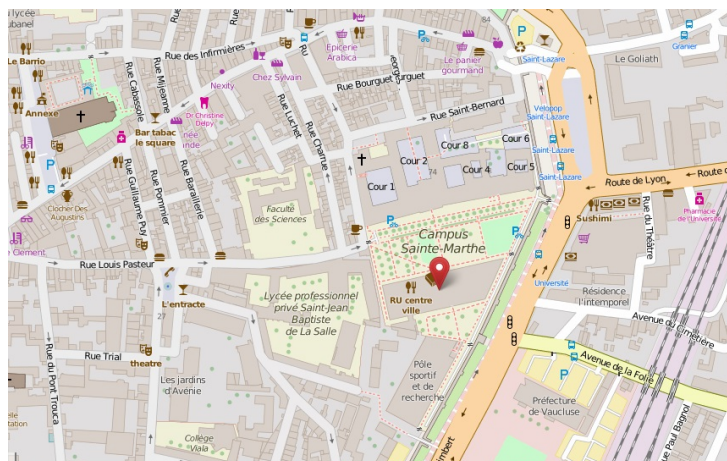


FIGURE 2. Localisation d'un point d'observation

```
<?xml version="1.0" encoding="UTF-8"?>
- <osm license="http://opendatacommons.org/licenses/odbl/1-0/"
  attribution="http://www.openstreetmap.org/copyright"
  copyright="OpenStreetMap and contributors" generator="CGImap 0.4.0 (8701
  thorn-02.openstreetmap.org)" version="0.6">
  <bounds maxlon="4.8056900" maxlat="43.9456000" minlon="4.8055300"
    minlat="43.9454300"/>
  - <node version="3" lon="4.8055510" lat="43.9454611" uid="538005"
    user="EddieJ" timestamp="2013-08-17T06:40:57Z" changeset="17379879"
    visible="true" id="2133489487">
    <tag v="parking_space" k="amenity"/>
    <tag v="1" k="capacity:disabled"/>
    <tag v="surface" k="parking"/>
    <tag v="disabled" k="parking_space"/>
    <tag v="survey 2013" k="source"/>
    <tag v="yes" k="wheelchair"/>
  </node>
  - <node version="1" lon="4.8056773" lat="43.9454860" uid="188930"
    user="J-Louis ZIMMERMANN" timestamp="2013-07-12T17:03:56Z"
    changeset="16928235" visible="true" id="2382554717">
    <tag v="waste_basket" k="amenity"/>
    <tag v="Mairie d'Avignon" k="operator"/>
    <tag v="survey 2013" k="source"/>
  </node>
  - <node version="2" lon="4.8121984" lat="43.9495394" uid="2117145"
    user="abdeldjalil" timestamp="2015-09-18T17:16:29Z"
    changeset="34107480" visible="true" id="2499310786">
    <tag v="Traiteur Garip" k="name"/>
    <tag v="deli" k="shop"/>
    <tag v="survey 2013" k="source"/>
    <tag v="no" k="wheelchair"/>
  </node>
  - <node version="4" lon="4.8119055" lat="43.9495545" uid="1068293"
    user="x0c0" timestamp="2014-05-04T15:01:37Z" changeset="22127250">
```

FIGURE 3. Exemple de fichier XML d'OSM sur Avignon

chaque requête soit court, si l'on veut une représentation efficace, d'où l'utilisation des bases de données pour stocker l'information environnementale.

```

F:\privacy\osm2pgsql\c\gwin-package-osm2pgsql -U postgres -d gis_paca -P 5432 -H localhost -s -S ./default.style "F:\privacy\osm2pgsql\gisData\paca.osm"
osm2pgsql SVN version 0.85.0 (64bit id space)

```

		id	lat	lon	id_node	tags
<input type="checkbox"/>	Modifier Copier Effacer	1	545669099	53579766	152546	{highway,crossing}
<input type="checkbox"/>	Modifier Copier Effacer	2	545829480	53527143	152590	{crossing,traffic_signals,highway,traffic_signals}

FIGURE 4. API OSM vers postgres et extrait de la table Nodes avec les Tags

La table des nœuds contient la *latitude* et la *longitude*. On peut donc appliquer les lois trigonométriques pour calculer la distance et l'orientation (Zdeb, 2010). On fait appel à l'API Google de géolocalisation inverse pour trouver les adresses des lieux. La table *Nodes* contient aussi le vecteur *Tags*. Ce dernier contient les détails introduits par les contributeurs d'OSM concernant chaque lieu (cf. Figure 4).

4.2. Calcul des distances entre le malvoyant et les lieux

Nous appliquons les lois de la trigonométrie dans l'espace. La formule *Haversine* nous permet de calculer la distance entre deux endroits en utilisant leur latitude et de leur longitude (Zdeb, 2010). Elle donne une distance à vol d'oiseau approximative, mais acceptable pour l'appréhension de l'espace géographique. Selon la formule *Haversine*, on calcule plusieurs paramètres a , puis c et la *distance*:

$$a = \sin^2(\delta\phi/2) + \cos\phi_1 \cdot \cos\phi_2 \cdot \sin^2(\delta\lambda/2) \quad (1)$$

$$c = 2 \cdot \arctan 2(\sqrt{a}, \sqrt{1-a}) \quad (2)$$

$$distance = R \cdot c \quad (3)$$

Avec

- $R = 6371000$ est le rayon de la terre en mètre ;
- ϕ_1 est latitude du malvoyant, (lat_1) en radian ;
- ϕ_2 est latitude du lieu (nœud) représenté sur OSM ; (lat_2) en Radian ;
- $\delta\phi = lat_2 - lat_1$ la différence des latitudes en Radian ;
- $\delta\lambda = lon_2 - lon_1$ la différence des longitudes en Radian ou (lon_2) est la longitude du lieu et (lon_1) est la longitude de le malvoyant.

On peut déduire la distance⁵ :

$$distance = \arccos(\sin(\phi_2) \cdot \sin(\phi_1) + \cos(\phi_2) \cdot \cos(\phi_1) \cdot \cos(\delta\lambda)) \cdot R \quad (4)$$

5. <http://www.movable-type.co.uk/scripts/latlong.html>

L'orientation du lieu (*bearing*) par rapport au malvoyant est calculée comme suit :

$$\theta = \arctan 2(\sin \delta\lambda \cdot \cos \phi_2, \cos \phi_1 \cdot \sin \phi_2 - \sin \phi_1 \cdot \cos \phi_2 \cdot \cos \delta\lambda) \quad (5)$$

Le résultat est un angle en radian entre $-\pi$ à π .

4.3. Recherche de tous les lieux dans un rayon donnée par le malvoyant

Supposant que le malvoyant cherche tous les lieux situés dans une distance d , les formules précédentes peuvent nous aider à calculer les distances en formulant les requêtes *sql*. En traduisant les latitudes et longitudes en distance relative d , nous pouvons ainsi trouver tous les lieux à l'intérieur du cercle dont le rayon est d .

$$r = d/R \quad (6)$$

où R est le rayon de la terre.

Les latitudes minimale et maximale à la distance d sont (Matuschek, 2010) :

$$lat_{min} = lat - r \quad (7)$$

$$lat_{max} = lat + r \quad (8)$$

Les longitudes minimale et maximale à la distance d sont:

$$lon_{min} = lon - \delta lon \quad (9)$$

$$lon_{max} = lon + \delta lon \quad (10)$$

où :

$$\delta lon = \arcsin(\sin(r) / \cos(lat)) \quad (11)$$

Le calcul des coordonnées GPS minimales et maximales pour d défini d'avance nous permet de formuler des requêtes de sélection vers la table des noeuds :

```
SELECT * FROM nodes WHERE (Lat >= lat_min AND Lat <= lat_max)
AND (Lon >= lon_min AND Lon <= lon_max)
```

4.4. Synthèse vocale et fonctionnalités du simulateur

Les caractéristiques recherchées du système sont les suivantes:

- Le moyen de communication principale est la voix ;
- Le système est *prudent* ; il arrive parfois que le système confonde les mots proche en consonance (ex: bus - base) ou interprète un bruit en un ordre ; le système demande la confirmation du choix du lieu, la sauvegarde des positions et le rayon de recherche ;

- Le système cherche l'*efficacité* ; pour que le processus de clarification sémantique (*grounding*) n'affecte l'efficacité du système en allongeant le temps d'interaction, on demande la confirmation de la tâche ;
- Le système est *flexible* ; le malvoyant peut changer sa position par ordre vocal, avant de chercher les lieux les plus proches de sa nouvelle position ;
- Le système est *multi-processus* ; pour rendre le système plus efficace, on crée deux processus de reconnaissance vocale, un pour les options du système, l'autre pour rechercher les lieux, le schéma de la figure 5 montre les attributs linguistique de chaque processus.

```
// To define list of options and list of search.independent each from other.
private Choices optionChoices = new Choices();
private Choices searchChoices = new Choices();

// Two speech recognition Engine, one for options the other to select and manipulate data from DB
private SpeechRecognitionEngine search_option_recEngine = new SpeechRecognitionEngine(CultureInfo.CurrentCulture);
private SpeechRecognitionEngine option_recEngine = new SpeechRecognitionEngine(CultureInfo.CurrentCulture);

// Speech synthesis, translate Text to Speech
SpeechSynthesizer speechSynth = new SpeechSynthesizer();
PromptBuilder promptBuild = new PromptBuilder();
```

FIGURE 5. création des commandes vocales sous C#

La synthèse vocale est la technique informatique de synthèse sonore qui permet de transformer un texte en parole artificielle. Le physicien théoricien britannique Stephen Hawking est l'une des personnes les plus connues utilisant ce genre de technique pour communiquer. Il faut noter que «*les documents numériques accessibles aux malvoyants par lecteur d'écran nécessitent que les images soient décrites*» (Allaire, 2012). Parmi les lecteurs de contenu d'écran sur le web, citons *TalkBack* de Firefox, *ChromeVox* de Chrome ou *VoiceOver* d'Apple.

La Figure 6 montre les deux processus STT (Speech To Text) que nous avons implémentés : à gauche, il concerne les options du système ; à droite, il concerne la recherche dans la base des données. Pour que le malvoyant passe d'une commande de catégorie (A) à une commande de catégorie (B), il doit activer le processus de reconnaissance correspondant. Voici un exemple de dialogue.

- Le malvoyant dit *Option Système* ; le simulateur répond *Options Système activées* ;
- Le malvoyant dit *Plus grand* ; le simulateur répond *Le rayon de recherche est de 600 mètres* ;
- Le malvoyant dit *Chercher dans la base* ; le simulateur répond *Recherche dans la base activée* ;
- Le malvoyant dit *Restaurant* ; le simulateur répond *Nous avons trouvé N restaurants* ;
- En revanche, si le malvoyant dit *Plus grand*, le simulateur ne répond pas car cet attribut n'appartient pas à *Chercher dans la base*.

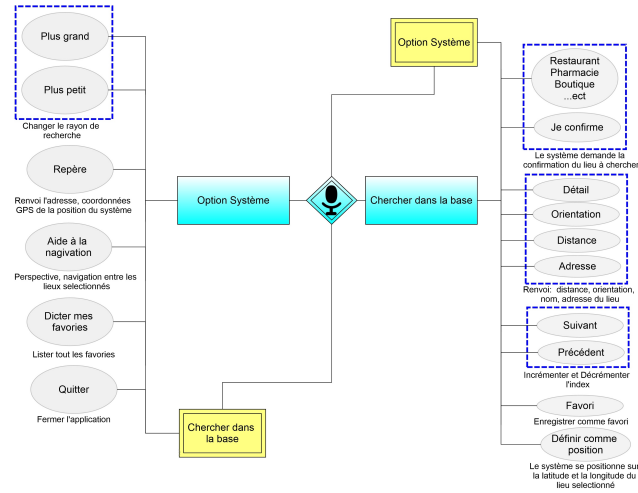


FIGURE 6. Modèle de dialogue Homme-Machine dédié aux malvoyants

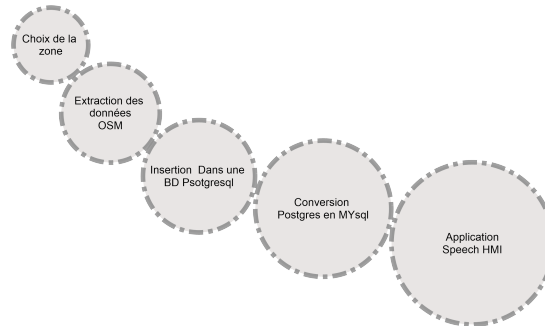
Pour résumer, le modèle de dialogue est simple. Le malvoyant peut choisir entre deux grandes fonctions :

Option système A ; les options sont : *augmenter ou diminuer le rayon de recherche, connaître son propre emplacement et lire la liste de ses lieux favoris.*

Chercher dans la base B ; la recherche dans la base exige une ontologie de recherche pour réduire les mots clés et réduire le temps de réponse. Il faut noter que les attributs d'OpenStreetMap sont en anglais et dans un langage formel. De plus, OSM permet aux contributeurs d'ajouter des attributs personnalisés. Donc une méthode sémantique de fouille dans la table des données est nécessaire pour extraire le maximum d'information et réduire suffisamment la liste des attributs vocaux du simulateur. Le processus de reconnaissance asynchrone du système (choix personnel) implique de définir la liste des mots utilisables.

4.5. Algorithme de transcription vocale

La figure 7 montre les cinq principales phases de construction de notre SIG vocal : *le choix de la zone géographique, l'extraction des données vectorielles, l'insertion des données dans une base Postgres, la conversion vers une base de données Mysql, l'exploitation des données par le simulateur vocal.* Le schéma de la figure 8 présente l'architecture et le fonctionnement du simulateur. Il montre la tâche de chaque commande, le défilement d'informations depuis la table des données *Mysql* jusqu'au malvoyant et la tâche de chaque commande vocale. Nous ne le détaillerons pas dans cet article.

FIGURE 7. *Cycles de développement de l'information*

4.6. Exemple de requêtes

Positionné à l'université d'Avignon, l'utilisateur envoie une requête vocale pour trouver des restaurants. La figure 9 montre les restaurants situés dans un rayon de 500 m. L'interface est organisée autour d'une boussole qui donne l'orientation. Il est également possible d'intégrer dans la requête une condition d'angle de recherche. À terme, celle-ci sera reliée à un capteur embarqué par le malvoyant pour qu'il puisse disposer des informations vocales lors de son déplacement réel.

5. Conclusion

Notre proposition permet potentiellement, en utilisant des données de type OSM au sein d'un SIG vocal, d'aider les malvoyants à se mouvoir dans leur environnement. Ce prototype sera testé prochainement avec des personnes d'associations de malvoyants. Les différences de perception et d'usage entre malvoyants et aveugles seront évaluées. En effet, l'intelligence artificielle a fait progresser l'interaction homme-machine, ce qui peut être bénéfique à la compensation du handicap chez certaines personnes à mobilité réduite. Le temps de réponse du simulateur, présenté dans cet article, est très court (moins de 2 secondes), mais on ne peut pas encore conclure avec certitude que la représentation sonore, ou plus précisément vocale, est de même efficacité qu'une représentation graphique. En effet, le nombre des attributs d'OpenStreetMap risque d'allonger le temps de réponse et aussi de rendre le système plus confus. La construction d'une ontologie pour la recherche d'informations à partir d'une base de données est ainsi probablement nécessaire pour remédier à ces problèmes. L'application de la synthèse et la reconnaissance vocales montre tout de même qu'il est possible de représenter les données vectorielles par la parole sans perdre le sens de l'information qu'elles portent, avec un système souple et adaptatif. En perspective, nous pensons placer le simulateur vocal au cœur

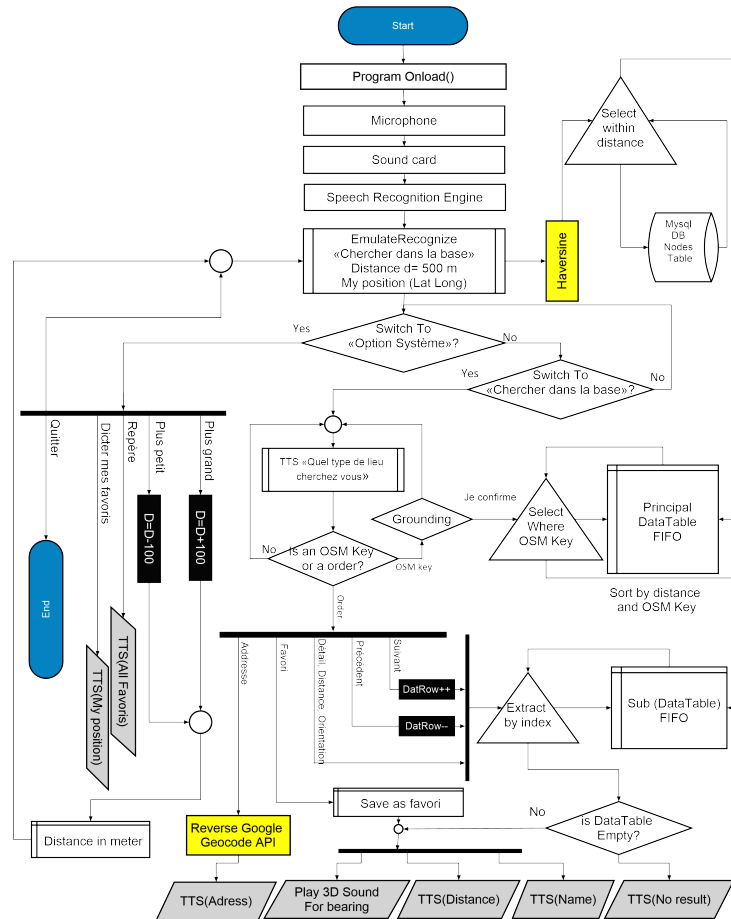


FIGURE 8. Structure de l'algorithme de transcription vocale

d'un système multimodal dont une entrée sera un capteur gyroscopique pour informer le malvoyant des lieux présents dans la direction de son déplacement.

Bibliographie

- Allaire C. (2012). *Informier les personnes aveugles ou malvoyantes*. Rapport technique. INPES.
- Baltenneck N. (2010). *Se mouvoir sans voir. Incidences de l'environnement urbain sur la perception, la représentation mentale et le stress lors du déplacement de la personne aveugle*. (Thèse de doctorat, Université Lumière Lyon 2)
- Bertin J. (1973). *Sémiologie graphique* (G. Villars, Ed.).

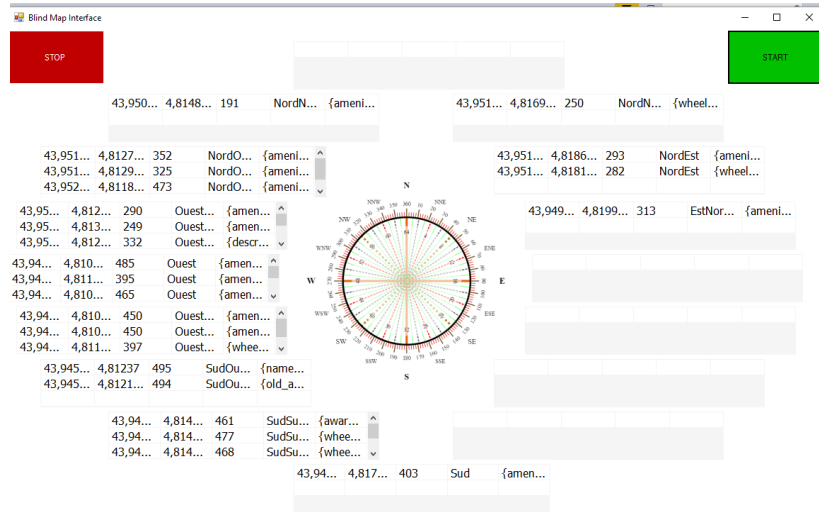


FIGURE 9. Recherche de restaurants dans un rayon de 500 m (localisation dans l'enceinte de l'Université d'Avignon)

- CFPSAA. (2010). *Les besoins des personnes déficientes visuelles*. Rapport technique. Confédération Française pour la Promotion Sociale des Aveugles et Amblyopes.
- Cha J. S., Lim D. K., Shin Y.-N. (2013). *Design and implementation of a voice based navigation for visually impaired persons*.
- Codescu M., Horsinka G., Kutz O., Mossakowski T., Rau R. (2011). DO-RAOM: activity-oriented search and navigation with openstreetmap. GeoSpatial Semantics: 4th International Conference, GeoS, Brest, France, may 2011.
- De Saussure F. (1997). *Cours de linguistique générale* (C. Bally, A. Sechehaye, Eds.).
- Gaver W. W. (1992). The affordances of media spaces for collaboration. Proceeding CSCW '92 Proceedings of the 1992 ACM conference on Computer-supported co-operative work Pages 17-24.
- Hatwell Y. (2006). Appréhender l'espace pour un enfant aveugle. *CAIRN.INFO*.
- Josselin D., Saidi A., Roussel D., Boularouk S., Bonin O., Altman E. *et al.* (2016). Sonorous cartography for sighted and blind people. Proceeding of AGILE'2016, Helsinki, 14-17 June, 2016. (6 pages, awarded best paper)
- Jouhaneau J. (2013). Perception de l'espace et immersion - perception auditive. *Techniques de l'ingénieur*.
- Kushwaha A., Kushwaha V. (2011). Location based services using android mobile operating system. *International Journal of Advances in Engineering & Technology*.
- Matuschek J. P. (2010). *Finding points within a distance of a latitude longitude using bounding coordinates*. Consulté sur <http://janmatuschek.de/LatitudeLongitudeBoundingCoordinates>

- Miura T., Ebihara Y., Sakajiri M., Ifukube T. (2011). Utilization of auditory perceptions of sounds and silent objects for orientation and mobility by visually-impaired people. IEEE International. (Institute of Gerontology, the University of Tokyo, Bunkyo-ku, Japan)
- Tenbrink T., Wiener J. M., Claramunt C. (Eds.). (2013). *Representing space in cognition: Interrelations of behaviour, language, and formal models*. Oxford University Press, Linguistics, 08 Exploration in Language and Space.
- Virole B. (2009). *Surdité et sciences humaines* (Chapitre 8, L'Harmattan éd.).
- Zdeb M. (2010). Driving distances and times using sas and google maps. SAS Global Forum 2010.